

Run on: July 6, 2004, 13:57:03 ; Search time 52.508 Seconds  
 (without alignments)  
 1382.059 Million cell updates/sec

Title: US-10-063-588-80  
 Perfect score: 1174  
 Sequence: 1 MASLGLQLVGYILGLLG.....PGQPPKVKSEFNSYSLTGYV 230

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : SPTREMBL\_25:\*  
 1: sp\_archaea:\*  
 2: sp\_bacteria:\*  
 3: sp\_fungi:\*  
 4: sp\_human:\*  
 5: sp\_invertebrate:\*  
 6: sp\_mammal:\*  
 7: sp\_mhc:\*  
 8: sp\_organelle:\*  
 9: sp\_phage:\*  
 10: sp\_plant:\*  
 11: sp\_rodent:\*  
 12: sp\_virus:\*  
 13: sp Vertebrate:\*  
 14: sp\_unclassified:\*  
 15: sp\_rvirus:\*  
 16: sp\_bacteriap:\*  
 17: sp\_archeap:\*

# SUMMARIES

Result		%	Query				
No.	Score	Match	Length	DB	ID	Description	
1	456	38.8	218	13	Q90XR2	Q90xr2 brachydanio	
2	452	38.5	213	13	Q7ZSZ2	Q7zsz2 xenopus lae	
3	448.5	38.2	219	11	Q810I9	Q810i9 mus musculu	
4	448	38.2	213	13	Q805G0	Q805g0 xenopus lae	
5	446	38.0	215	13	Q90XR8	Q90xr8 brachydanio	
6	444.5	37.9	209	13	Q90XR9	Q90xr9 brachydanio	
7	438	37.3	210	13	Q90XQ8	Q90xq8 brachydanio	
8	432.5	36.8	214	13	Q90XQ9	Q90xq9 xenopus lae	
9	430	36.6	214	13	Q9DE12	Q9de12 xenopus lae	
10	428.5	36.5	214	13	Q98SR2	Q98sr2 gallus gall	
11	421	35.9	214	13	Q90XR0	Q90xr0 brachydanio	

12	419	35.7	214	13	Q7T2P4	Q7t2p4 brachydanio
13	417.5	35.6	210	13	Q90XR4	Q90xr4 brachydanio
14	413.5	35.2	206	13	Q90XS1	Q90xs1 brachydanio
15	399.5	34.0	211	4	Q7Z4X9	Q7z4x9 homo sapien
16	396	33.7	211			

Run on: July 6, 2004, 13:58:13 ; Search time 19.9678 Seconds  
 (without alignments)  
 1107.984 Million cell updates/sec

Title: US-10-063-588-80  
 Perfect score: 1174  
 Sequence: 1 MASLGLQLVGYILGLLGLLG.....PGQPPKVKSEFNSYSLTGYV 230

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : PIR\_78:\*  
 1: pir1:\*  
 2: pir2:\*  
 3: pir3:\*  
 4: pir4:\*

# SUMMARIES

Result			%		Query		DB	ID	Description
	No.	Score	Match	Length					
1	423.5	36.1	280	2	A39484			androgen-withdrawa	
2	100	8.5	400	2	AG2866			MFS permease [drug	
3	100	8.5	400	2	D97643			probable efflux pr	
4	99.5	8.5	330	2	C69648			2-keto-3-deoxygluc	
5	99.5	8.5	463	2	F90285			metabolite transpo	
6	97.5	8.3	484	2	C75609			amino acid ABC tra	
7	95.5	8.1	458	2	H71657			NADH2 dehydrogenas	
8	93	7.9	523	2	T11916			NADH2 dehydrogenas	
9	91	7.8	160	2	JN0503			peripheral myelin	
10	90.5	7.7	487	2	T19237			hypothetical prote	
11	90.5	7.7	617	2	F71019			hypothetical prote	
12	88.5	7.5	429	2	A97241			permease [imported	
13	88.5	7.5	521	2	A99549			amino acid permeas	
14	88.5	7.5	593	2	H71283			conserved hypothet	
15	88	7.5	160	2	A41144			growth arrest-rela	

Run on: July 6, 2004, 13:56:23 ; Search time 14.791 Seconds  
(without alignments)  
809.691 Million cell updates/sec

Title: US-10-063-588-80  
Perfect score: 1174  
Sequence: 1 MASLGLQLVGYYILGLLGLLG.....PGQPPKVKSEFNLSYSLTGYV 230

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result		%					
No.	Score	Query	Match	Length	DB	ID	Description
1	1174	100.0	230	1	CLD2_HUMAN	P57739	homo sapien
2	1119	95.3	230	1	CLD2_CANFA	Q95km6	canis famil
3	1096	93.4	230	1	CLD2_MOUSE	O88552	mus musculu
4	571	48.6	239	1	CLDE_HUMAN	O95500	homo sapien
5	561.5	47.8	239	1	CLDE_MOUSE	Q9z0s3	mus musculu
6	459.5	39.1	208	1	CLDY_BRARE	Q9yh91	brachydanio
7	450.5	38.4	218	1	CLD3_CANFA	Q95km5	canis famil
8	450	38.3	220	1	CLD3_HUMAN	O15551	homo sapien
9	446	38.0	210	1	CLD4_MOUSE	O35054	mus musculu
10	443.5	37.8	209	1	CLD4_CERAE	O19005	cercopithec

XX  
PS Claim 1; Page 12 (Disclosure); 42pp; Chinese.

XX  
CC The present sequence represents human SP82 protein, which has cancer-  
CC suppressing activity. The present invention also describes a method for  
CC the preparation of the protein by recombination, and the application of  
CC the protein in treating diseases such as cancer

XX  
SQ Sequence 230 AA;

Query Match	100.0%;	Score 1174;	DB 5;	Length 230;	
Best Local Similarity	100.0%;	Pred. No. 7.3e-117;			
Matches 230;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	

Qy	1	MASLGLQLVGYYLGLLGLLGLTLVAMLLPSWKTSSYVQASIVTAVGFSKGLWMECATHSTG	60
Db	1	MASLGLQLVGYYLGLLGLLGLTLVAMLLPSWKTSSYVQASIVTAVGFSKGLWMECATHSTG	60
Qy	61	ITQCDIYSTLLGLPADIAQAQAMMVTSSAIISSLACIISVVGMRCTVFCQESRAKDRVAVA	120
Db	61	ITQCDIYSTLLGLPADIAQAQAMMVTSSAIISSLACIISVVGMRCTVFCQESRAKDRVAVA	120
Qy	121	GGVFFILGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII	180
Db	121	GGVFFILGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII	180
Qy	181	LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGOPPKVKSEFNSSYSLTGYV	230
Db	181	LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGOPPKVKSEFNSSYSLTGYV	230

Search completed: July 6, 2004, 14:00:34  
Job time : 76.2154 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

```
Run on:      July  6, 2004, 13:58:48 ; Search time 63.6013 Seconds
              (without alignments)
              1125.689 Million cell updates/sec
```

```
Title:          US-10-063-588-80
Perfect score:  1174
Sequence:       1 MASLGLQLVGYILGLLG.....PGOPPKVKSEFNSYSLTGYV 230
```

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1276540 seqs, 311283816 residues

Total number of hits satisfying chosen parameters: 1276540

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	1174	100.0	230	9	US-09-998-598-2590	Sequence 2590, Ap
2	1174	100.0	230	10	US-09-888-257A-6	Sequence 6, Appli
3	1174	100.0	230	10	US-09-946-374-134	Sequence 134, App
4	1174	100.0	230	11	US-09-833-245-1256	Sequence 1256, Ap
5	1174	100.0	230	12	US-10-147-493-492	Sequence 492, App
6	1174	100.0	230	12	US-10-145-127-492	Sequence 492, App
7	1174	100.0	230	12	US-10-160-503-492	Sequence 492, App
8	1174	100.0	230	12	US-10-143-118-492	Sequence 492, App
9	1174	100.0	230	12	US-10-144-993-492	Sequence 492, App
10	1174	100.0	230	12	US-10-158-787-492	Sequence 492, App
11	1174	100.0	230	12	US-10-081-056-192	Sequence 192, App
12	1174	100.0	230	12	US-10-063-745-80	Sequence 80, Appl
13	1174	100.0	230	12	US-10-140-024-492	Sequence 492, App
14	1174	100.0	230	12	US-10-063-512-80	Sequence 80, Appl
15	1174	100.0	230	12	US-10-063-513-80	Sequence 80, Appl
16	1174	100.0	230	12	US-10-063-515-80	Sequence 80, Appl
17	1174	100.0	230	12	US-10-063-549-80	Sequence 80, Appl
18	1174	100.0	230	12	US-10-063-569-80	Sequence 80, Appl
19	1174	100.0	230	12	US-10-063-551-80	Sequence 80, Appl
20	1174	100.0	230	12	US-10-140-808-492	Sequence 492, App
21	1174	100.0	230	12	US-10-006-485A-134	Sequence 134, App
22	1174	100.0	230	12	US-10-013-907A-134	Sequence 134, App
23	1174	100.0	230	12	US-10-015-499A-134	Sequence 134, App

Run on: July 6, 2004, 13:55:23 ; Search time 73.2154 Seconds  
 (without alignments)  
 887.599 Million cell updates/sec

Title: US-10-063-588-80  
 Perfect score: 1174  
 Sequence: 1 MASLGLQLVGYYILGLLGLLG.....PGQPPKVKSEFNSYSLTGYV 230

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
 1: geneseqp1980s:\*  
 2: geneseqp1990s:\*  
 3: geneseqp2000s:\*  
 4: geneseqp2001s:\*  
 5: geneseqp2002s:\*  
 6: geneseqp2003as:\*  
 7: geneseqp2003bs:\*  
 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	1174	100.0	230	2 AAY36181	Aay36181 Human sec
2	1174	100.0	230	3 AAY84609	Aay84609 A human m
3	1174	100.0	230	3 AAY99378	Aay99378 Human PRO
4	1174	100.0	230	4 AAB66127	Aab66127 Protein o
5	1174	100.0	230	4 AAE04207	Aae04207 Human gen
6	1174	100.0	230	4 AAU12417	Aau12417 Human PRO
7	1174	100.0	230	4 AAU09178	Aau09178 Human PRO
8	1174	100.0	230	4 AAB87565	Aab87565 Human PRO
9	1174	100.0	230	4 AAB88342	Aab88342 Human mem
10	1174	100.0	230	5 ABP67991	Abp67991 Human col
11	1174	100.0	230	5 ABG95890	Abg95890 Human sec
12	1174	100.0	230	5 ABB84912	Abb84912 Human PRO
13	1174	100.0	230	5 AAU76534	Aau76534 Tumour-as
14	1174	100.0	230	5 ABG64507	Abg64507 Human alb
15	1174	100.0	230	5 ABB04707	Abb04707 Human SP8
16	1174	100.0	230	5 ABB95518	Abb95518 Human ang

17	1174	100.0	230	6	ABO17861	Abo17861	Novel	hum
18	1174	100.0	230	6	ABU81115	Abu81115	Human	PRO
19	1174	100.0	230	6	ABU90915	Abu90915	Novel	hum
20	1174	100.0	230	6	ABO33974	Abo33974	Human	sec
21	1174	100.0	230	6	ABU71991	Abu71991	Novel	hum
22	1174	100.0	230	6	ABU66815	Abu66815	Human	PRO
23	1174	100.0	230	6	ABU71545	Abu71545	Human	sec
24	1174	100.0	230	6	ABR47946	Abr47946	Human	sec
25	1174	100.0	230	6	ABU59896	Abu59896	Novel	sec
26	1174	100.0	230	6	ABU72326	Abu72326	Human	PRO
27	1174	100.0	230	6	ABU90999	Abu90999	Human	PRO
28	1174	100.0	230	6	ABO27320	Abo27320	Human	sec
29	1174	100.0	230	6	ABO25086	Abo25086	Human	sec
30	1174	100.0	230	6	ABR00178	Abr00178	Human	gen
31	1174	100.0	230	6	ABU92515	Abu92515	Human	sec
32	1174	100.0	230	6	ABU81185	Abu81185	Human	sec
33	1174	100.0	230	6	ABO53299	Abo53299	Novel	hum
34	1174	100.0	230	6	ABU67091	Abu67091	Human	sec
35	1174	100.0	230	6	ABU98302	Abu98302	Novel	hum
36	1174	100.0	230	6	ABU89307	Abu89307	Novel	hum
37	1174	100.0	230	6	ABU82514	Abu82514	Novel	hum
38	1174	100.0	230	6	ABU96478	Abu96478	Human	PRO
39	1174	100.0	230	6	ABU72148	Abu72148	Human	PRO
40	1174	100.0	230	6	ADA46011	Ada46011	Novel	hum
41	1174	100.0	230	6	ADA76442	Ada76442	Human	PRO
42	1174	100.0	230	6	ADB17137	Adb17137	Human	tra
43	1174	100.0	230	6	ADA19092	Ada19092	Human	PRO
44	1174	100.0	230	6	ADA61715	Ada61715	Homo sapi	
45	1174	100.0	230	6	ADB19500	Adb19500	Novel	hum

# ALIGNMENTS

RESULT 1  
AAY36181



Run on: July 6, 2004, 13:55:23 ; Search time 73.2154 Seconds  
 (without alignments)  
 887.599 Million cell updates/sec

Title: US-10-063-588-80  
 Perfect score: 1174  
 Sequence: 1 MASLGLQLVG YILG LLLG.....PGQPPKVKSEFNSYSLTGYV 230

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*  
 1: geneseqp1980s:\*  
 2: geneseqp1990s:\*  
 3: geneseqp2000s:\*  
 4: geneseqp2001s:\*  
 5: geneseqp2002s:\*  
 6: geneseqp2003as:\*  
 7: geneseqp2003bs:\*  
 8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result	Score	% Query Match	Length	DB	ID	Description
No.						
1	1174	100.0	230	2	AAAY36181	Aay36181 Human sec
2	1174	100.0	230	3	AAAY84609	Aay84609 A human m
3	1174	100.0	230	3	AAAY99378	Aay99378 Human PRO
4	1174	100.0	230	4	AAB66127	Aab66127 Protein o
5	1174	100.0	230	4	AAE04207	Aae04207 Human gen
6	1174	100.0	230	4	AAU12417	Aau12417 Human PRO
7	1174	100.0	230	4	AAU09178	Aau09178 Human PRO
8	1174	100.0	230	4	AAB87565	Aab87565 Human PRO
9	1174	100.0	230	4	AAB88342	Aab88342 Human mem
10	1174	100.0	230	5	ABP67991	Abp67991 Human col
11	1174	100.0	230	5	ABG95890	Abg95890 Human sec
12	1174	100.0	230	5	ABB84912	Abb84912 Human PRO
13	1174	100.0	230	5	AAU76534	Aau76534 Tumour-as
14	1174	100.0	230	5	ABG64507	Abg64507 Human alb
15	1174	100.0	230	5	ABB04707	Abb04707 Human SP8
16	1174	100.0	230	5	ABB95518	Abb95518 Human ang

Run on: July 6, 2004, 10:17:26 ; Search time 6305 Seconds  
(without alignments)  
10139.722 Million cell updates/sec

Title: US-10-063-588-79  
Perfect score: 1475  
Sequence: 1 gagagaagtcagcctggcag.....ctgggacatttaaaaaaata 1475

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*

35: em\_htg\_rod:\*  
 36: em\_htg\_mam:\*  
 37: em\_htg\_vrt:\*  
 38: em\_sy:\*  
 39: em\_htgo\_hum:\*  
 40: em\_htgo\_mus:\*  
 41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	1475	100.0	1475	6	AX092348	AX092348 Sequence
2	1475	100.0	1475	6	AX299996	AX299996 Sequence
3	1475	100.0	1475	6	AX395213	AX395213 Sequence
4	1475	100.0	1475	6	AX454606	AX454606 Sequence
5	1475	100.0	1475	6	AX464358	AX464358 Sequence
6	1475	100.0	1475	6	AX491084	AX491084 Sequence
7	1475	100.0	1475	6	AX697065	AX697065 Sequence
8	1475	100.0	1475	9	AY358474	AY358474 Homo sapi
9	1475	100.0	2959	9	AK075405	AK075405 Homo sapi
10	1475	100.0	109465	9	AL158821	AL158821 Human DNA
11	1473.4	99.9	1506	9	BC014424	BC014424 Homo sapi
12	1473.4	99.9	1918	9	AF177340	AF177340 Homo sapi
13	1472	99.8	2742	6	BD237562	BD237562 Membrane-
14	1470.2	99.7	2863	6	AX136129	AX136129 Sequence
15						

Run on: July 6, 2004, 10:17:25 ; Search time 4776.8 Seconds  
(without alignments)  
9220.961 Million cell updates/sec

Title: US-10-063-588-79  
Perfect score: 1475  
Sequence: 1 gagagaagtcagcctggcag.....ctgggacatttaaaaaaata 1475

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_htc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_htc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vrt:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_mam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rod:\*
- 26: em\_gss\_phg:\*
- 27: em\_gss\_vrl:\*
- 28: gb\_gss1:\*
- 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result			Query		DB	ID	Description
No.	Score	% Match	Length				
1	817.8	55.4	2829	11	AK004990	AK004990	Mus muscu
2	731	49.6	822	12	BG386074	BG386074	602455248
3	693	47.0	693	29	AY400251	AY400251	Homo sapi
4	689.8	46.8	945	10	BE513091	BE513091	601171545
5	676.4	45.9	693	29	AY400252	AY400252	Pan trogl
6	643	43.6	680	10	BE304667	BE304667	601105783
7	637.4	43.2	728	12	BG325755	BG325755	602424466
8	617	41.8	978	12	BG164062	BG164062	602341087
9	614.4	41.7	786	28	BZ215522	BZ215522	CH230-416
10	606	41.1	643	12	BG328625	BG328625	602427889
11	582.2	39.5	589	13	BU071883	BU071883	im35d11.y
c 12	577	39.1	676	10	AW385836	AW385836	CM2-LT004
c 13	575.6	39.0	630	10	AW837727	AW837727	CM1-LT004
c 14	555.4	37.7	623	10	AW837724	AW837724	CM1-LT004
c 15	555	37.6	555	13	BU073403	BU073403	im35d11.x
16	552.6	37.5	777				

Run on: July 6, 2004, 10:17:25 ; Search time 702.749 Seconds  
 (without alignments)  
 8916.548 Million cell updates/sec

Title: US-10-063-588-79  
 Perfect score: 1475  
 Sequence: 1 gagagaagtcagcctggcag.....ctgggacatttaaaaaaata 1475

Scoring table: IDENTITY\_NUC  
 Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
 1: geneseqn1980s:\*  
 2: geneseqn1990s:\*  
 3: geneseqn2000s:\*  
 4: geneseqn2001as:\*  
 5: geneseqn2001bs:\*  
 6: geneseqn2002s:\*  
 7: geneseqn2003as:\*  
 8: geneseqn2003bs:\*  
 9: geneseqn2003cs:\*  
 10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result		%					
No.	Score	Query Match	Length	DB	ID	Description	
1	1475	100.0	1475	3	AAA37060	Aaa37060 Human PRO	
2	1475	100.0	1475	4	AAF54296	Aaf54296 DNA encod	
3	1475	100.0	1475	4	AAS21489	Aas21489 Human cDN	
4	1475	100.0	1475	4	AAS15360	Aas15360 cDNA enco	
5	1475	100.0	1475	4	AAF92097	Aaf92097 Human PRO	
6	1475	100.0	1475	6	ABS74417	Abs74417 Human cDN	
7	1475	100.0	1475	6	ABL88167	Abl88167 Human PRO	
8	1475	100.0	1475	6	ABK11089	Abk11089 cDNA enco	
9	1475	100.0	1475	6	ABL95656	Abl95656 Human ang	
10	1475	100.0	1475	7	ACD24098	Acd24098 Novel hum	
11	1475	100.0	1475	7	ACA67239	Aca67239 cDNA enco	
12	1475	100.0	1475	7	ACA91203	Aca91203 Novel hum	
13	1475	100.0	1475	7	ACD81580	Acd81580 Human cDN	
14	1475	100.0	1475	7	ACA60402	Aca60402 Novel hum	

15	1475	100.0	1475	7	ACA03848	Aca03848	cDNA	enco
16	1475	100.0	1475	7	ACA58849	Aca58849	cDNA	enco
17	1475	100.0	1475	7	ABX89386	Abx89386	DNA	encod
18	1475	100.0	1475	7	ACA64025	Aca64025	cDNA	enco
19	1475	100.0	1475	7	ACA91289	Aca91289	cDNA	enco
20	1475	100.0	1475	7	ACD45188	Acd45188	Human	sec
21	1475	100.0	1475	7	ACD42040	Acd42040	Human	sec
22	1475	100.0	1475	7	ACA93736	Aca93736	Human	cDN
23	1475	100.0	1475	7	ACA67310			

AF177340  
 LOCUS AF177340 1918 bp mRNA linear PRI 03-OCT-2000  
 DEFINITION Homo sapiens clone SP82 claudin 2 mRNA, complete cds.  
 ACCESSION AF177340  
 VERSION AF177340.1 GI:10503979  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1918)  
 AUTHORS Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,  
 Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,  
 Yu,J. and Han,L.H.  
 TITLE Novel human cDNA clone with function of inhibiting cancer cell  
 growth  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1918)  
 AUTHORS Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,  
 Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,  
 Yu,J. and Han,L.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-AUG-1999) National Laboratory For Oncogenes & Related  
 Genes, Shanghai Cancer Institute, 25/Ln 2200 Xie Tu Road, Shanghai  
 200032, P.R. China  
 FEATURES Location/Qualifiers  
 source 1. 1918  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="SP82"  
 CDS 520..1212  
 /note="similar to Mus musculus claudin 2"  
 /codon\_start=1  
 /product="claudin 2"  
 /protein\_id="AAG17984.1"  
 /db\_xref="GI:10503980"  
 /translation="MASLGLQLVGYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAV  
 GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAIISSLACIISVVGM  
 RCTVFCQESRAKDRVAVAGGVFFILGLLGFIPVAWNHLHGILRDFYSPLVPDSMKFEI  
 GEALYLGIISSLFSLIAGIILCFSCSSQRNRSNYDAYQAQPLATRSSPRPGQPPKVK  
 SEFNYSYSLTGYV"

# ORIGIN

Query Match 99.9%; Score 1473.4; DB 9; Length 1918;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1474; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GAGAGAAGTCAGCCTGGCAGAGAGACTCTGAAATGAGGGATTAGAGGTGTTCAAGGAGCA 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 399 GAGAGAAGTCAGCCTGGCAGAGAGACTCTGAAATGAGGGATTAGAGGTGTTCAAGGAGCA 458  
 QY 61 AGAGCTTCAGCCTGAAGACAAGGGAGCAGTCCCTGAAGACGCTTCTACTGAGAGGTCTGC 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 459 AGAGCTTCAGCCTGAAGACAAGGGAGCAGTCCCTGAAGACGCTTCTACTGAGAGGTCTGC 518  
 QY 121 CATGGCCTCTCTTGGCCTCCAACCTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGG 180



Db	519	 CATGGCCTCTCTTGGCCTCCAACCTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGG	578
Qy	181	CACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCAT	240
Db	579	 CACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCAT	638
Qy	241	TGTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGG	300
Db	639	 TGTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGG	698
Qy	301	CATCACCAGTGACATCTATAGCACCCTTCTGGGCCTGCCCCTGACATCCAGGCTGC	360
Db	699	 CATCACCAGTGACATCTATAGCACCCTTCTGGGCCTGCCCCTGACATCCAGGCTGC	758
Qy	361	CCAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGT	420
Db	759	 CCAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGT	818
Qy	421	GGGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGC	480
Db	819	 GGGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGC	878
Qy	481	AGGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCTGTTGCCTGGAATCT	540
Db	879	 AGGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCTGTTGCCTGGAATCT	938
Qy	541	TCATGGGATCCTACGGGACTTCTACTCACCCTGGTGCTGACAGCATGAAATTTGAGAT	600
Db	939	 TCATGGGATCCTACGGGACTTCTACTCACCCTGGTGCTGACAGCATGAAATTTGAGAT	998
Qy	601	TGGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATCAT	660
Db	999	 TGGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATCAT	1058
Qy	661	CCTCTGCTTTTTCCTGCTCATCCCAGAGAAATCGCTCCAACCTACTACGATGCCTACCAAGC	720
Db	1059	 CCTCTGCTTTTTCCTGCTCATCCCAGAGAAATCGCTCCAACCTACTACGATGCCTACCAAGC	1118
Qy	721	CCAACCTCTTGCCACAAGGAGCTCTCCAAGGCCTGGTCAACCTCCCAAAGTCAAGAGTGA	780
Db	1119	 CCAACCTCTTGCCACAAGGAGCTCTCCAAGGCCTGGTCAACCTCCCAAAGTCAAGAGTGA	1178
Qy	781	GTTCAATTCCTACAGCCTGACAGGGTATGTGTGAAGAACCAGGGGCCAGAGCTGGGGGGT	840
Db	1179	 GTTCAATTCCTACAGCCTGACAGGGTATGTGTGAAGAACCAGGGGCCAGAGCTGGGGGGT	1238
Qy	841	GGCTGGGTCTGTGAAAAACAGTGGACAGCACCCCGAGGGCCACAGGTGAGGGACACTACC	900
Db	1239	 GGCTGGGTCTGTGAAAAACAGTGGACAGCACCCCGAGGGCCACAGGTGAGGGACACTACC	1298
Qy	901	ACTGGATCGTGTCAGAAGGTGCTGCTGAGGATAGACTGACTTTGGCCATTGGATTGAGCA	960
Db	1299	 ACTGGATCGTGTCAGAAGGTGCTGCTGAGGATAGACTGACTTTGGCCATTGGATTGAGCA	1358
Qy	961	AAGGCAGAAATGGGGGCTAGTGTAACAGCATGCAGGTTGAATTGCCAAGGATGCTCGCCA	1020

Db	1359	AAGGCAGAAATGGGGGCTAGTGTAACAGCATGCAGGTTGAATTGCCAAGGATGCTCGCCA	1418
Qy	1021	TGCCAGCCTTTCTGTTTTCTCACCTTGCTGCTCCCCTGCCCTAAGTCCCCAACCCCTCAA	1080
Db	1419	TGCCAGCCTTTCTGTTTTCTCACCTTGCTGCTCCCCTGCCCTAAGTCCCCAACCCCTCAA	1478
Qy	1081	CTTGAAACCCCATTCCTTAAGCCAGGACTCAGAGGATCCCTTTGCCCTCTGGTTTACCT	1140
Db	1479	CTTGAAACCCCATTCCTTAAGCCAGGACTCAGAGGATCCCTTTGCCCTCTGGTTTACCT	1538
Qy	1141	GGGACTCCATCCCCAAACCCACTAATCACATCCCACTGACTGACCCTCTGTGATCAAAGA	1200
Db	1539	GGGACTCCATCCCCAAACCCACTAATCACATCCCACTGACTGACCCTCTGTGATCAAAGA	1598
Qy	1201	CCCTCTCTCTGGCTGAGGTTGGCTCTTAGCTCATTGCTGGGGATGGGAAGGAGAAGCAGT	1260
Db	1599	CCCTCTCTCTGGCTGAGGTTGGCTCTTAGCTCATTGCTGGGGATGGGAAGGAGAAGCAGT	1658
Qy	1261	GGCTTTTGTGGGCATTGCTCTAACCTACTTCTCAAGCTTCCCTCCAAAGAACTGATTGG	1320
Db	1659	GGCTTTTGTGGGCATTGCTCTAACCTACTTCTCAAGCTTCCCTCCAAAGAACTGATTGG	1718
Qy	1321	CCCTGGAACCTCCATCCCACTCTTGTTATGACTCCACAGTGTCCAGACTAATTTGTGCAT	1380
Db	1719	CCCTGGAACCTCCATCCCACTCTTGTTATGACTCCACAGTGTCCAGACTAATTTGTGCAT	1778
Qy	1381	GAACTGAAATAAAACCATCCTACGGTATCCAGGGAACAGAAAGCAGGATGCAGGATGGGA	1440
Db	1779	GAACTGAAATAAAACCATCCTACGGTATCCAGGGAACAGAAAGCAGGATGCAGGATGGGA	1838
Qy	1441	GGACAGGAAGGCAGCCTGGGACATTTAAAAAAATA	1475
Db	1839	GGACAGGAAGGCAGCCTGGGACATTTAAAAAAA	1873

ID CLD2\_HUMAN STANDARD; PRT; 230 AA.  
 AC P57739;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Claudin-2 (UNQ705/PRO1356).  
 GN CLDN2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Epithelium;  
 RA Reinecker H.-C., Sakaguchi T., Golden H.M.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,  
 RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,  
 RA Yu J., Han L.H.;  
 RT "Novel human cDNA clone with function of inhibiting cancer cell  
 RT growth.";  
 RL Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22887296; PubMed=12975309;  
 RA Clark H.F., Gurney A.L., Abaya E., Baker K. et al.  
 RT "The secreted protein discovery initiative (SPDI), a large-scale  
 RT effort to identify novel human secreted and transmembrane proteins: a  
 RT bioinformatics assessment.";  
 RL Genome Res. 13:2265-2270(2003).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Heath P.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Colon, and Kidney;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

```

RT  "Generation and initial analysis of more than 15,000 full-length
RT  human and mouse cDNA sequences.";
RL  Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC  -!- FUNCTION: Component of tight junction (TJ) strands.
CC  -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC  -!- SIMILARITY: Belongs to the claudin family.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AF250558; AAF98151.1; -.
DR  EMBL; AF177340; AAG17984.1; -.
DR  EMBL; AY358474; AAQ88838.1; -.
DR  EMBL; AL158821; CAD23055.1; -.
DR  EMBL; BC015252; AAH15252.1; -.
DR  EMBL; BC014424; AAH14424.1; -.
DR  Genew; HGNC:2041; CLDN2.
DR  InterPro; IPR006187; Claudin.
DR  InterPro; IPR005411; Claudin2.
DR  InterPro; IPR006188; Claudin_reg.
DR  InterPro; IPR004031; PMP22_Claudin.
DR  Pfam; PF00822; PMP22_Claudin; 1.
DR  PRINTS; PR01077; CLAUDIN.
DR  PRINTS; PR01589; CLAUDIN2.
DR  PROSITE; PS01346; CLAUDIN; 1.
KW  Tight junction; Transmembrane.
FT  TRANSMEM      8      28      POTENTIAL.
FT  TRANSMEM     82     102      POTENTIAL.
FT  TRANSMEM    117     137      POTENTIAL.
FT  TRANSMEM    163     183      POTENTIAL.
SQ  SEQUENCE    230 AA;  24548 MW;  52CA642D4A62B70D CRC64;

```

Qy	1	MASLGLQLVGYYILGLLGLLGLTLVAMLLPSWKTS	SSYVGASIVTAVGFSKGLWMECATHSTG	60
Db	1	MASLGLQLVGYYILGLLGLLGLTLVAMLLPSWKTS	SSYVGASIVTAVGFSKGLWMECATHSTG	60
Qy	61	ITQCDIYSTLLGLPADIQAAQAMMVTSSA	ISSLACIISVVGMRCTVFCQESRAKDRVAVA	120
Db	61	ITQCDIYSTLLGLPADIQAAQAMMVTSSA	ISSLACIISVVGMRCTVFCQESRAKDRVAVA	120
Qy	121	GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII	180	
Db	121	GGVFFILGGLLGFIPVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII	180	
Qy	181	LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGOPPKVKSEFN	SYSLTGYV	230
Db	181	LCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGOPPKVKSEFN	SYSLTGYV	230

Search  for

Limits Preview/Index History Clipboard Details

Range: from  to  Features: ☐ SNP ☐ CDD ☒ MGC ☐ HPRD ☐ STS

☐ 1: [P57739](#). Reports Claudin-2 (UNQ705...[gi:12229749]

[BLink](#), [Domains](#),  
[Links](#)

LOCUS P57739 230 aa linear PRI 01-MAY-2005  
 DEFINITION Claudin-2 (UNQ705/PRO1356) (SP82).  
 ACCESSION P57739  
 VERSION P57739 GI:12229749  
 DBSOURCE swissprot: locus CLD2\_HUMAN, accession [P57739](#);  
 class: standard.  
 created: Oct 16, 2001.  
 sequence updated: Oct 16, 2001.  
 annotation updated: May 1, 2005.  
 xrefs: [AF250558.1](#), [AAF98151.1](#), [AF177340.1](#), [AAG17984.1](#), [AY358474.1](#),  
[AAQ88838.1](#), [AK075371.1](#), [BAC11575.1](#), [AL158821.16](#), [CAD23055.1](#),  
[BC015252.1](#), [AAH15252.1](#), [BC014424.1](#), [AAH14424.1](#), [BC071747.1](#),  
[AAH71747.1](#)  
 xrefs (non-sequence databases): EnsemblENSG00000165376,  
 GenewHGNC:2041, H-InvDBHIX0016971, MIM [300520](#), GO0016021,  
 GO0005923, GO0042802, GO0016338, InterProIPR006187,  
 InterProIPR005411, InterProIPR006188, InterProIPR004031,  
 PfamPF00822, PRINTSPR01077, PRINTSPR01589, PROSITEPS01346  
 KEYWORDS Tight junction; Transmembrane.  
 SOURCE Homo sapiens (human)  
 ORGANISM [Homo sapiens](#)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Hominidae; Homo.  
 REFERENCE 1 (residues 1 to 230)  
 AUTHORS Sakaguchi,T., Gu,X., Golden,H.M., Suh,E., Rhoads,D.B. and  
 Reinecker,H.C.  
 TITLE Cloning of the human claudin-2 5'-flanking region revealed a  
 TATA-less promoter with conserved binding sites in mouse and human  
 for caudal-related homeodomain proteins and hepatocyte nuclear  
 factor-lalpha  
 JOURNAL J. Biol. Chem. 277 (24), 21361-21370 (2002)  
 PUBMED [11934881](#)  
 REMARK NUCLEOTIDE SEQUENCE.  
 TISSUE=Colon epithelium  
 REFERENCE 2 (residues 1 to 230)  
 AUTHORS Gu,J.R., Wan,D.F., Zhao,X.T., Zhou,X.M., Jiang,H.Q., Zhang,P.P.,  
 Qin,W.X., Huang,Y., Qiu,X.K., Qian,L.F., He,L.P., Li,H.N., Yu,Y.,  
 Yu,J. and Han,L.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (??-AUG-1999)  
 REMARK NUCLEOTIDE SEQUENCE.  
 REFERENCE 3 (residues 1 to 230)  
 AUTHORS Clark,H.F., Gurney,A.L., Abaya,E., Baker,K., Baldwin,D., Brush,J.,  
 Chen,J., Chow,B., Chui,C., Crowley,C., Currell,B., Deuel,B.,  
 Dowd,P., Eaton,D., Foster,J., Grimaldi,C., Gu,Q., Hass,P.E.,  
 Heldens,S., Huang,A., Kim,H.S., Klimowski,L., Jin,Y., Johnson,S.,  
 Lee,J., Lewis,L., Liao,D., Mark,M., Robbie,E., Sanchez,C.,  
 Schoenfeld,J., Seshagiri,S., Simmons,L., Singh,J., Smith,V.,  
 Stinson,J., Vagts,A., Vandlen,R., Watanabe,C., Wieand,D., Woods,K.,  
 Xie,M.H., Yansura,D., Yi,S., Yu,G., Yuan,J., Zhang,M., Zhang,Z.,



	Goddard,A., Wood,W.I., Godowski,P. and Gray,A.
TITLE	The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: a bioinformatics assessment
JOURNAL	Genome Res. 13 (10), 2265-2270 (2003)
PUBMED	<u>12975309</u>
REMARK	NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
REFERENCE	4 (residues 1 to 230)
AUTHORS	Ota,T., Suzuki,Y., Nishikawa,T., Otsuki,T., Sugiyama,T., Irie,R., Wakamatsu,A., Hayashi,K., Sato,H., Nagai,K., Kimura,K., Makita,H., Sekine,M., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Ishii,S., Yamamoto,J., Saito,K., Kawai,Y., Isono,Y., Nakamura,Y., Nagahari,K., Murakami,K., Yasuda,T., Iwayanagi,T., Wagatsuma,M., Shiratori,A., Sudo,H., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Kanda,K., Yokoi,T., Furuya,T., Kikkawa,E., Omura,Y., Abe,K., Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M., Ninomiya,K., Ishibashi,T., Yamashita,H., Murakawa,K., Fujimori,K., Tanai,H., Kimata,M., Watanabe,M., Hiraoka,S., Chiba,Y., Ishida,S., Ono,Y., Takiguchi,S., Watanabe,S., Yosida,M., Hotuta,T., Kusano,J., Kanehori,K., Takahashi-Fujii,A., Hara,H., Tanase,T.O., Nomura,Y., Togiya,S., Komai,F., Hara,R., Takeuchi,K., Arita,M., Imose,N., Musashino,K., Yuuki,H., Oshima,A., Sasaki,N., Aotsuka,S., Yoshikawa,Y., Matsunawa,H., Ichihara,T., Shiohata,N., Sano,S., Moriya,S., Momiyama,H., Satoh,N., Takami,S., Terashima,Y., Suzuki,O., Nakagawa,S., Senoh,A., Mizoguchi,H., Goto,Y., Shimizu,F., Wakebe,H., Hishigaki,H., Watanabe,T., Sugiyama,A., Takemoto,M., Kawakami,B., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyama,M., Tashiro,H., Tanigami,A., Fujiwara,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hirao,M., Ohmori,Y., Kawabata,A., Hikiiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T., Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Sasaki,M., Togashi,T., Oyama,M., Hata,H., Watanabe,M., Komatsu,T., Mizushima-Sugano,J., Satoh,T., Shirai,Y., Takahashi,Y., Nakagawa,K., Okumura,K., Nagase,T., Nomura,N., Kikuchi,H., Masuho,Y., Yamashita,R., Nakai,K., Yada,T., Nakamura,Y., Ohara,O., Isogai,T. and Sugano,S.
TITLE	Complete sequencing and characterization of 21,243 full-length human cDNAs
JOURNAL	Nat. Genet. 36 (1), 40-45 (2004)
PUBMED	<u>14702039</u>
REMARK	NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA]. TISSUE=Teratocarcinoma
REFERENCE	5 (residues 1 to 230)
AUTHORS	Ross,M.T., Grafham,D.V., Coffey,A.J., Scherer,S., McLay,K., Muzny,D., Platzer,M., Howell,G.R., Burrows,C., Bird,C.P., Frankish,A., Lovell,F.L., Howe,K.L., Ashurst,J.L., Fulton,R.S., Sudbrak,R., Wen,G., Jones,M.C., Hurles,M.E., Andrews,T.D., Scott,C.E., Searle,S., Ramser,J., Whittaker,A., Deadman,R., Carter,N.P., Hunt,S.E., Chen,R., Cree,A., Gunaratne,P., Havlak,P., Hodgson,A., Metzker,M.L., Richards,S., Scott,G., Steffen,D., Sodergren,E., Wheeler,D.A., Worley,K.C., Ainscough,R., Ambrose,K.D., Ansari-Lari,M.A., Aradhya,S., Ashwell,R.I., Babbage,A.K., Bagguley,C.L., Ballabio,A., Banerjee,R., Barker,G.E., Barlow,K.F., Barrett,I.P., Bates,K.N., Beare,D.M., Beasley,H., Beasley,O., Beck,A., Bethel,G., Blechschmidt,K., Brady,N., Bray-Allen,S., Bridgeman,A.M., Brown,A.J., Brown,M.J., Bonnini,D., Bruford,E.A., Buhay,C., Burch,P., Burford,D., Burgess,J., Burrill,W., Burton,J., Bye,J.M., Carder,C., Carrel,L., Chako,J., Chapman,J.C., Chavez,D., Chen,E., Chen,G., Chen,Y., Chen,Z., Chinault,C., Ciccodicola,A., Clark,S.Y., Clarke,G., Clee,C.M., Clegg,S., Clerc-Blankenburg,K., Clifford,K., Copley,V., Cole,C.G., Conquer,J.S., Corby,N., Connor,R.E., David,R., Davies,J., Davis,C., Davis,J., Delgado,O., Deshazo,D., Dhami,P., Ding,Y., Dinh,H., Dodsworth,S., Draper,H., Dugan-Rocha,S., Dunham,A., Dunn,M., Durbin,K.J., Dutta,I., Eades,T., Ellwood,M., Emery-Cohen,A.,

Errington, H., Evans, K.L., Faulkner, L., Francis, F., Frankland, J., Fraser, A.E., Galgoczy, P., Gilbert, J., Gill, R., Glockner, G., Gregory, S.G., Gribble, S., Griffiths, C., Grocock, R., Gu, Y., Gwilliam, R., Hamilton, C., Hart, E.A., Hawes, A., Heath, P.D., Heitmann, K., Hennig, S., Hernandez, J., Hinzmann, B., Ho, S., Hoffs, M., Howden, P.J., Huckle, E.J., Hume, J., Hunt, P.J., Hunt, A.R., Isherwood, J., Jacob, L., Johnson, D., Jones, S., de Jong, P.J., Joseph, S.S., Keenan, S., Kelly, S., Kershaw, J.K., Khan, Z., Kioschis, P., Klages, S., Knights, A.J., Kosiura, A., Kovar-Smith, C., Laird, G.K., Langford, C., Lawlor, S., Leversha, M., Lewis, L., Liu, W., Lloyd, C., Lloyd, D.M., Loulseged, H., Loveland, J.E., Lovell, J.D., Lozado, R., Lu, J., Lyne, R., Ma, J., Maheshwari, M., Matthews, L.H., McDowall, J., McLaren, S., McMurray, A., Meidl, P., Meitinger, T., Milne, S., Miner, G., Mistry, S.L., Morgan, M., Morris, S., Muller, I., Mullikin, J.C., Nguyen, N., Nordsiek, G., Nyakatura, G., O'Dell, C.N., Okwuonu, G., Palmer, S., Pandian, R., Parker, D., Parrish, J., Pasternak, S., Patel, D., Pearce, A.V., Pearson, D.M., Pelan, S.E., Perez, L., Porter, K.M., Ramsey, Y., Reichwald, K., Rhodes, S., Ridler, K.A., Schlessinger, D., Schueler, M.G., Sehra, H.K., Shaw-Smith, C., Shen, H., Sheridan, E.M., Shownkeen, R., Skuce, C.D., Smith, M.L., Sotheran, E.C., Steingruber, H.E., Steward, C.A., Storey, R., Swann, R.M., Swarbreck, D., Tabor, P.E., Taudien, S., Taylor, T., Teague, B., Thomas, K., Thorpe, A., Timms, K., Tracey, A., Trevanion, S., Tromans, A.C., d'Urso, M., Verduzco, D., Villasana, D., Waldron, L., Wall, M., Wang, Q., Warren, J., Warry, G.L., Wei, X., West, A., Whitehead, S.L., Whiteley, M.N., Wilkinson, J.E., Willey, D.L., Williams, G., Williams, L., Williamson, A., Williamson, H., Wilming, L., Woodmansey, R.L., Wray, P.W., Yen, J., Zhang, J., Zhou, J., Zoghbi, H., Zorilla, S., Buck, D., Reinhardt, R., Poustka, A., Rosenthal, A., Lehrach, H., Meindl, A., Minx, P.J., Hillier, L.W., Willard, H.F., Wilson, R.K., Waterston, R.H., Rice, C.M., Vaudin, M., Coulson, A., Nelson, D.L., Weinstock, G., Sulston, J.E., Durbin, R., Hubbard, T., Gibbs, R.A., Beck, S., Rogers, J. and Bentley, D.R.

TITLE	The DNA sequence of the human X chromosome
JOURNAL	Nature 434 (7031), 325-337 (2005)
PUBMED	<u>15772651</u>
REMARK	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
REFERENCE	6 (residues 1 to 230)
AUTHORS	Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus, D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
CONSRTM	Mammalian Gene Collection Program Team
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
PUBMED	<u>12477932</u>
REMARK	NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
	TISSUE=Colon, Kidney, and Skin
COMMENT	[FUNCTION] Plays a major role in tight junction-specific obliteration of the intercellular space, through

calcium-independent cell-adhesion activity (By similarity).  
 [SUBUNIT] Can form homo- and heteropolymers with other CLDN.  
 Homopolymers interact with CLDN3, but not CLDN1, homopolymers.  
 Directly interacts with TJP1/ZO-1, TJP2/ZO-2 and TJP3/ZO-3 (By  
 similarity).  
 [SUBCELLULAR LOCATION] Integral membrane protein. Tight junctions  
 (By similarity).  
 [SIMILARITY] Belongs to the claudin family.

FEATURES	Location/Qualifiers
source	1..230 /organism="Homo sapiens" /db_xref="taxon:9606"
gene	1..230 /gene="CLDN2"
Protein	1..230 /gene="CLDN2" /product="Claudin-2"
Region	1..7 /gene="CLDN2" /region_name="Domain" /note="Cytoplasmic (Potential)." /evidence=not_experimental
Region	8..28 /gene="CLDN2" /region_name="Transmembrane region" /note="Potential." /evidence=not_experimental
Region	29..81 /gene="CLDN2" /region_name="Domain" /note="Extracellular (Potential)." /evidence=not_experimental
Region	82..102 /gene="CLDN2" /region_name="Transmembrane region" /note="Potential." /evidence=not_experimental
Region	103..116 /gene="CLDN2" /region_name="Domain" /note="Cytoplasmic (Potential)." /evidence=not_experimental
Region	117..137 /gene="CLDN2" /region_name="Transmembrane region" /note="Potential." /evidence=not_experimental
Region	138..162 /gene="CLDN2" /region_name="Domain" /note="Extracellular (Potential)." /evidence=not_experimental
Region	163..183 /gene="CLDN2" /region_name="Transmembrane region" /note="Potential." /evidence=not_experimental
Region	184..230 /gene="CLDN2" /region_name="Domain" /note="Cytoplasmic (Potential)." /evidence=not_experimental
Site	229..230 /gene="CLDN2" /site_type="unclassified" /note="Interactions with TJP1, TJP2 and TJP3 (By



similarity)."
/evidence=not\_experimental

ORIGIN

1 maslglqlvg yilglglglg tlvamllpsw ktssyvgasi vtavgfsgkl wmecathstg
61 itqcdiystl lglpadiqaa qammvtssai sslaciisvv gmrctvfcqe srakdrvava
121 ggvvffilggl lgfipvawnl hgilrdfysp lvpdsmkfei gealylgiis slfsliagii
181 lcfscssqrn rsnyydayqa qplatrsspr pgqppkvkse fnsysltgyv

//

Disclaimer | Write to the Help Desk
NCBI | NLM | NIH

Feb 9 2005 14:51:10

[PubMed](#)
[Nucleotide](#)
[Protein](#)
[Genome](#)
[Structure](#)
[PMC](#)
[Taxonomy](#)
[OMIM](#)
[Books](#)

Search  for

[Limits](#)
[Preview/Index](#)
[History](#)
[Clipboard](#)
[Details](#)





Range: from  to  ☐ Reverse complemented strand Features: ☐ SNP ☐ CDD

☒ MGC ☐ HPRD ☐ STS

☐ 1: [NM\\_020384](#). Reports Homo sapiens clau...[gi:38455423]

[Links](#)

LOCUS NM\_020384 2959 bp mRNA linear PRI 23-APR-2005

DEFINITION Homo sapiens claudin 2 (CLDN2), mRNA.

ACCESSION NM\_020384

VERSION NM\_020384.2 GI:38455423

KEYWORDS .

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
Hominidae; Homo.

REFERENCE 1 (bases 1 to 2959)

AUTHORS Mankertz,J., Hillenbrand,B., Tavalali,S., Huber,O., Fromm,M. and Schulzke,J.D.

TITLE Functional crosstalk between Wnt signaling and Cdx-related transcriptional activation in the regulation of the claudin-2 promoter activity

JOURNAL Biochem. Biophys. Res. Commun. 314 (4), 1001-1007 (2004)

PUBMED [14751232](#)

REMARK GeneRIF: role for the paracellular barrier function by opening pores for small cations

REFERENCE 2 (bases 1 to 2959)

AUTHORS Colegio,O.R., Van Itallie,C., Rahner,C. and Anderson,J.M.

TITLE Claudin extracellular domains determine paracellular charge selectivity and resistance but not tight junction fibril architecture

JOURNAL Am. J. Physiol., Cell Physiol. 284 (6), C1346-C1354 (2003)

PUBMED [12700140](#)

REMARK GeneRIF: Results support a model in which claudins 2 and 4 create paracellular channels and the first extracellular domain is sufficient to determine both paracellular charge selectivity and transepithelial electrical resistance.

REFERENCE 3 (bases 1 to 2959)

AUTHORS Gonzalez-Mariscal,L., Betanzos,A., Nava,P. and Jaramillo,B.E.

TITLE Tight junction proteins

JOURNAL Prog. Biophys. Mol. Biol. 81 (1), 1-44 (2003)

PUBMED [12475568](#)

REMARK Review article

REFERENCE 4 (bases 1 to 2959)

AUTHORS Tsukita,S. and Furuse,M.

TITLE Claudin-based barrier in simple and stratified cellular sheets

JOURNAL Curr. Opin. Cell Biol. 14 (5), 531-536 (2002)

PUBMED [12231346](#)

REMARK Review article

REFERENCE 5 (bases 1 to 2959)

AUTHORS Sakaguchi,T., Gu,X., Golden,H.M., Suh,E., Rhoads,D.B. and Reinecker,H.C.

TITLE Cloning of the human claudin-2 5'-flanking region revealed a TATA-less promoter with conserved binding sites in mouse and human for caudal-related homeodomain proteins and hepatocyte nuclear factor-1alpha

JOURNAL J. Biol. Chem. 277 (24), 21361-21370 (2002)

PUBMED 11934881  
 REMARK GeneRIF: Cloning of the human claudin-2 5'-flanking region revealed a TATA-less promoter with conserved binding sites in mouse and human for caudal-related homeodomain proteins and hepatocyte nuclear factor-1alpha.

REFERENCE 6 (bases 1 to 2959)  
 AUTHORS Tsukita,S., Furuse,M. and Itoh,M.  
 TITLE Multifunctional strands in tight junctions  
 JOURNAL Nat. Rev. Mol. Cell Biol. 2 (4), 285-293 (2001)  
 PUBMED 11283726  
 REMARK Review article

REFERENCE 7 (bases 1 to 2959)  
 AUTHORS Heiskala,M., Peterson,P.A. and Yang,Y.  
 TITLE The roles of claudin superfamily proteins in paracellular transport  
 JOURNAL Traffic 2 (2), 93-98 (2001)  
 PUBMED 11247307  
 REMARK Review article

REFERENCE 8 (bases 1 to 2959)  
 AUTHORS Kniesel,U. and Wolburg,H.  
 TITLE Tight junctions of the blood-brain barrier  
 JOURNAL Cell. Mol. Neurobiol. 20 (1), 57-76 (2000)  
 PUBMED 10690502  
 REMARK Review article

REFERENCE 9 (bases 1 to 2959)  
 AUTHORS Itoh,M., Furuse,M., Morita,K., Kubota,K., Saitou,M. and Tsukita,S.  
 TITLE Direct binding of three tight junction-associated MAGUKs, ZO-1, ZO-2, and ZO-3, with the COOH termini of claudins  
 JOURNAL J. Cell Biol. 147 (6), 1351-1363 (1999)  
 PUBMED 10601346

REFERENCE 10 (bases 1 to 2959)  
 AUTHORS Morita,K., Furuse,M., Fujimoto,K. and Tsukita,S.  
 TITLE Claudin multigene family encoding four-transmembrane domain protein components of tight junction strands  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (2), 511-516 (1999)  
 PUBMED 9892664

REFERENCE 11 (bases 1 to 2959)  
 AUTHORS Furuse,M., Fujita,K., Hiiiragi,T., Fujimoto,K. and Tsukita,S.  
 TITLE Claudin-1 and -2: novel integral membrane proteins localizing at tight junctions with no sequence similarity to occludin  
 JOURNAL J. Cell Biol. 141 (7), 1539-1550 (1998)  
 PUBMED 9647647

COMMENT VALIDATED REFSEQ: This record has undergone preliminary review of the sequence, but has not yet been subject to final review. The reference sequence was derived from AK075405.1.  
 On Nov 20, 2003 this sequence version replaced gi:9966780.

Summary: Members of the claudin protein family, such as CLDN2, are expressed in an organ-specific manner and regulate the tissue-specific physiologic properties of tight junctions (Sakaguchi et al., 2002).[supplied by OMIM].

FEATURES

source	Location/Qualifiers
	1..2959
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/chromosome="X"
	/map="Xq22.3-23"
gene	1..2959
	/gene="CLDN2"
	/db_xref="GeneID:9075"
	/db_xref="HPRD:HPRD_06471"
	/db_xref="MIM:300520"
CDS	327..1019
	/gene="CLDN2"
	/note="go_component: membrane [goid <u>0016020</u> ] [evidence IEA];

go\_component: tight junction [goid 0005923] [evidence  
 ISS];  
 go\_component: integral to membrane [goid 0016021]  
 [evidence ISS];  
 go\_function: protein self binding [goid 0042802] [evidence  
 ISS];  
 go\_function: structural molecule activity [goid 0005198]  
 [evidence IEA];  
 go\_process: calcium-independent cell-cell adhesion [goid  
0016338] [evidence ISS]"  
 /codon\_start=1  
 /product="claudin 2"  
 /protein\_id="NP\_065117.1"  
 /db\_xref="GI:9966781"  
 /db\_xref="CCDS:CCDS14524.1"  
 /db\_xref="GeneID:9075"  
 /db\_xref="HPRD:HPRD\_06471"  
 /db\_xref="MIM:300520"  
 /translation="MASLGLQLVGYILGLLGLLGLTLVAMLLPSWKTSSYVGASIVTAV  
 GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMTSSAIISSLACIISVVG  
 RCTVFCQESRAKDRVAVAGGVFFILGGLLGFIPVAWNHLHGILRDFYSPLVPDSMKFEI  
 GEALYLGIISSLFSLIAGIILCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVK  
 SEFNYSYSLTGYV"

# ORIGIN

```

1  agatggattt  tgcaaagctg  tgggttaacga  ttagaaatcc  tttatcacct  cagcccgtgg
61  ccccttgtag  ttgcgtcccc  tccctcagga  tccctttctc  cctctccagg  ggcattctcc
121 cctccaaggc  tctgcaaaga  actgccctgt  cttctagatg  ccttcttgag  gctgcttggt
181 gccacccaca  gacacttgta  aggaggagag  aagtcagcct  ggcagagaga  ctctgaaatg
241 agggattaga  ggtgttcaag  gagcaagagc  ttcagcctga  agacaaggga  gcagtccttg
301 aagacgcttc  tactgagagg  tctgccatgg  cctctcttgg  cctccaactt  gtgggctaca
361 tcctaggcct  tctggggcct  ttgggcacac  tggttgccat  gctgctcccc  agctggaaaa
421 caagtcttta  tgctcgggtg  agcattgtga  cagcagttgg  cttctccaag  ggcctctgga
481 tggaatgtgc  cacacacagc  acaggcatca  cccagtgtga  catctatagc  acccttctgg
541 gcctgcccgc  tgacatccag  gctgcccagg  ccatgatggt  gacatccagt  gcaatctcct
601 ccctggcctg  cattatctct  gtggtgggca  tgagatgcac  agtcttctgc  caggaatccc
661 gagccaaaga  cagagtggcg  gtagcaggtg  gagtcttttt  catccttgga  ggcctcctgg
721 gattcattcc  tgttgccctg  aatcttcatg  ggatcctacg  ggacttctac  tcaccactgg
781 tgccctgacg  catgaaatct  gagattggag  aggctcttta  cttgggcatt  atttcttccc
841 tgttctccct  gatagctgga  atcatectct  gcttttccct  ctcattcccag  agaaatcgct
901 ccaactacta  cgatgcctac  caagcccaac  ctcttgccac  aaggagctct  ccaaggcctg
961 gtcaacctcc  caaagtcaag  agtgagttca  attcctacag  cctgacaggg  tatgtgtgaa
1021 gaaccagggg  ccagagctgg  ggggtggctg  ggtctgtgaa  aaacagtgga  cagcaccctg
1081 agggccacag  gtgagggaca  ctaccactgg  atcgtgtcag  aaggtgctgc  tgaggataga
1141 ctgacttttg  ccattggatt  gagcaaaggc  agaaatgggg  gctagtgtaa  cagcatgcag
1201 gttgaattgc  caaggatgct  cgccatgcca  gcctttctgt  tttcctcacc  ttgctgctcc
1261 cctgccctaa  gtcccccaac  ctcaacttga  aaccccatte  ccttaagcca  ggactcagag
1321 gatccctttg  cctcttggtt  tacctgggac  tccatcccca  aaccactaa  tcacatccca
1381 ctgactgacc  ctctgtgatc  aaagaccctc  tctctggctg  aggttggtct  ttagctcatt
1441 gctggggatg  ggaaggagaa  gcagtggtct  ttgtgggcat  tgctctaacc  tacttctcaa
1501 gcttccctcc  aaagaaactg  attggccctg  gaacctccat  cccactcttg  ttatgactcc
1561 acagtgtcca  gactaatttg  tgcataaact  gaaataaaac  catcctacgg  tatccaggga
1621 acagaaagca  ggatgcagga  tgggaggaca  ggaaggcagc  ctgggacatt  taaaaaata
1681 aaaatgaaaa  aaaaaccag  aacccatttc  tcagggcact  ttccagaatt  ctctcatatt
1741 tgtgggctgg  gatcaagcct  gcagcttgag  gaaagcacia  ggaaaggaaa  gaagatctgg
1801 tggaaagctc  aggtggcagc  ggactctgac  tccactgagg  aactgcctca  gaagctgcga
1861 tcacaacttt  ggctgaagcc  cctgcctcac  tctagggcac  ctgacctggc  ctcttgctta
1921 aaccacaagg  ctaagggcta  tagacaatgg  tttccttagg  aacagtaaac  cagtttttct
1981 agggatggcc  cttggctggg  ggatgacagt  gtgggagctg  tgggggtact  aggaagacac
2041 cattccttga  cggtgtctaa  gaagccaggt  ggatgtgtgt  ggtggctcca  gtgggtgttt
2101 ctactctgcc  agtgagaggc  agccccctag  aaactcttca  ggcgtaatgg  aaaatcagct
2161 caaatgagat  caggccccc  cagggtccac  ccacagagca  ctacagagcc  tctgaaagac
2221 catagacca  agcgagcccc  ttcagattcc  cccactgtcc  atcggaagat  gctccagagt
2281 ggctagaggg  catctaaggg  ctccagcatg  gcatatccat  gccacgggtg  ctgtgtccat
2341 gatctgagtg  atagctgcac  tgctgcctgg  gattgcagct  gaggtgggag  tggagaatgg
2401 ttcccaggaa  gacagttcca  cctctaagg  ccgaaaatgt  tccctttacc  ctggagtggg
2461 agtgaggggt  catacaccaa  aggtattttc  cctcaccagt  ctaggcatga  ctggcttctg

```

2521 aaaaattcca gcacacctcc tcgaacctca ttgtcagcag agagggccca tctgttgtct  
2581 gtaacatgcc tttcacatgt ccaccttctt gccatgttcc agctgctctc ccaacctgga  
2641 aggccgtctc cccttagcca agtcctcctc aggcttgagg aacttcctca gcgtcacctc  
2701 cttcattgag ccttctctga tcaactccatc cctctcctac ccctccctcc cccaaccctc  
2761 aatgtataaa ttgcttcttg atgcttagca ttcacaattt ttgattgatc gttatttgtg  
2821 tgtgtgtgtc cgatctcaca agtatattgt aaacccttcg gtgggtgggg gccatatact  
2881 agacctctct gtatcccca gactatctgt aacagtgccca ggcacacagt aggtgatcaa  
2941 taaacacttg ttgattgag

//

[Disclaimer](#) | [Write to the Help Desk](#)  
[NCBI](#) | [NLM](#) | [NIH](#)

Feb 9 2005 14:31:10

LOCUS BE513091 945 bp mRNA linear EST 07-AUG-2000  
 DEFINITION 601171545F1 NIH\_MGC\_15 Homo sapiens cDNA clone IMAGE:3545171 5', mRNA sequence.  
 ACCESSION BE513091  
 VERSION BE513091.1 GI:9720302  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 945)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: NIH Intramural Sequencing Center  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: LLCM241 row: i column: 12  
 High quality sequence start: 36  
 High quality sequence stop: 782.

FEATURES Location/Qualifiers  
 source 1. .945  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3545171"  
 /tissue\_type="adenocarcinoma cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_15"  
 /note="Organ: colon; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)"

#### ORIGIN

Query Match 46.8%; Score 689.8; DB 10; Length 945;  
 Best Local Similarity 94.1%; Pred. No. 7.9e-190;  
 Matches 761; Conservative 0; Mismatches 42; Indels 6; Gaps 4;

Qy 1 GAGAGAAGTCAGCCTGGCAGAGAGACTCTGAAATGAGGGATTAGAGGTGTTCAAGGAGCA 60  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 66 GAGAGAAGTCAGCCTGGCAGAGAGACTCTGAAATGAGGGATTAGAGGTGTTCAAGGAGCA 125  
 Qy 61 AGAGCTTCAGCCTGAAGACAAGGGAGCAGTCCCTGAAGACGCTTCTACTGAGAGGTCTGC 120  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 126 AGAGCTTCAGCCTGAAGACAAGGGAGCAGTCCCTGAAGACGCTTCTACTGAGAGGTCTGC 185  
 Qy 121 CATGGCCTCTCTTGGCCTCCAACCTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGG 180



Db	186	 CATGGCCTCTCTTGGCCTCCAACCTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGG	245
Qy	181	CACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCAT	240
Db	246	 CACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCAT	305
Qy	241	TGTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGG	300
Db	306	 TGTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGG	365
Qy	301	CATCACCAGTGACATCTATAGCACCTTCTGGGCCTGCCCGCTGACATCCAGGCTGC	360
Db	366	 CATCACCAGTGACATCTATAGCACCTTCTGGGCCTGCCCGCTGACATCCAGGCTGC	425
Qy	361	CCAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGT	420
Db	426	 CCAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGT	485
Qy	421	GGGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGC	480
Db	486	 GGGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGC	545
Qy	481	AGGTGGAGTCTTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCTGTTGCCTGGAATCT	540
Db	546	 AGGTGGAGTC-TTTTTCATCCTTGGAGGCCTCCTGGGATTCATTCTGTTGCCTGGAATCT	604
Qy	541	TCATGGGATCCTACGGGACTTCTACTCACCAGTGGTGCTGACAGCATGAAATTTGA--G	598
Db	605	 TCATGGGATCCTACGGGACTTCTACTCACCAGTGGTGCTGACAGCATGAAATTTGAAGA	664
Qy	599	ATTGGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATC	658
Db	665	 TTGGCAGAGGCTCTTTACTTGGGCTTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATC	724
Qy	659	A-TCCTCTGCTTTTCCTGCTCATCCCAGAGAAATCGCTCCAACCTACTACGATGCCTACCA	717
Db	725	 ATTCCTCTGCTTTTCCTGCTCATCCCAGAGAAATCGGTCCAACCTACTACGATGCCTACCA	784
Qy	718	AGCCCAACCTCTTGCCACAAGGAGCTCTCCAAGGCCTGGTCAACCTCCCAAAGTCAAGAG	777
Db	785	 AGGCGCCACCTCTTGCCACAGGAGCTCTCCAGGGCTTGGTCACCTTCCCAAGGCAAGAG	844
Qy	778	TGAGTTCAATTCCTACAGCCTGACAGGGT	806
Db	845	 T--GGTTCATTCCTACAGCCTGACAGGGT	871

ID AAB88342 standard; protein; 230 AA.  
 DT 23-MAY-2001 (first entry)  
 DE Human membrane or secretory protein clone PSEC0059.  
 XX  
 KW Human; secretory protein; membrane protein; vaccine; gene therapy;  
 KW rheumatoid arthritis; diabetes.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1067182-A2.  
 XX  
 PD 10-JAN-2001.  
 XX  
 PF 07-JUL-2000; 2000EP-00114090.  
 XX  
 PR 08-JUL-1999; 99JP-00194179.  
 PR 11-JAN-2000; 2000JP-00118775.  
 PR 02-MAY-2000; 2000JP-00183766.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Kawai Y, Sugiyama T, Hayashi K;  
 XX  
 DR WPI; 2001-093989/11.  
 DR N-PSDB; AAF93769.  
 XX  
 PT Nucleic acids encoding secretory proteins/membrane proteins, useful in  
 PT gene therapy or as candidate target molecules in drug development.  
 XX  
 PS Claim 1; SEQ ID NO 52; 609pp + Sequence Listing; English.  
 XX  
 CC This invention relates to nucleic acid sequences AAF93744 - AAF93916  
 CC which encode human secretory or membrane proteins represented by AAB88317  
 CC - AAB88419. Included in the invention are primers AAF93917 - AAF94295 and  
 CC AAF62232 - AAF62235 which are used to isolate the cDNA sequences of the  
 CC invention. The invention also includes methods for the production of  
 CC antibodies directed against the proteins, and cDNA sequences, which can  
 CC be used in vaccines. The polynucleotide sequences can be used in gene  
 CC therapy. The polynucleotide sequences and the proteins they encode may be  
 CC used in the prevention, treatment and diagnosis of diseases associated  
 CC with inappropriate secretory protein/membrane protein expression. The  
 CC nucleic acids and complementary sequences may also be used as DNA probes  
 CC in diagnostic assays (e.g. polymerase chain reactions (PCR)) to detect  
 CC and quantitate the presence of similar nucleic acid sequences in samples.  
 CC They may also be used to study the expression and function of secretory  
 CC proteins/membrane polypeptides and their role in metabolism. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC against them and in assays to identify modulators (agonists and  
 CC antagonists) of expression and activity. The antibodies and antagonists  
 CC may also be used as therapeutic agents to down regulate expression and  
 CC activity. The antibodies may also be used as diagnostic agents for  
 CC detecting the presence of the polypeptides in samples (e.g. by enzyme  
 CC linked immunosorbant assay (ELISA). Examples of diseases which may be  
 CC treated include rheumatoid arthritis and diabetes  
 XX  
 SQ Sequence 230 AA;



Query Match 100.0%; Score 1174; DB 4; Length 230;  
Best Local Similarity 100.0%; Pred. No. 7.3e-117;  
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MASLGLQLVGYILGLLGLLGLTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60
        |||
Db      1 MASLGLQLVGYILGLLGLLGLTLVAMLLPSWKTSSYVGASIVTAVGFSKGLWMECATHSTG 60

Qy     61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAIISSLACIISVVGMRCTVFCQESRAKDRVAVA 120
        |||
Db     61 ITQCDIYSTLLGLPADIQAAQAMMVTSSAIISSLACIISVVGMRCTVFCQESRAKDRVAVA 120

Qy    121 GGVFFILGGLLGFI PVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180
        |||
Db    121 GGVFFILGGLLGFI PVAWNLHGILRDFYSPLVPDSMKFEIGEALYLGIISSLFSLIAGII 180

Qy    181 LCFSCSSQRNRSNYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
        |||
Db    181 LCFSCSSQRNRSNYDAYQAQPLATRSSPRPGQPPKVKSEFNSYSLTGYV 230
```

RESULT 10

ABP67991

ID ABP67991 standard; protein; 230 AA.

XX

AC ABP67991;

XX

DT 13-DEC-2002 (first entry)

XX

DE Human colon cancer related polypeptide SEQ ID NO 2590.

XX

KW Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine.

XX

OS Homo sapiens.

XX

PN WO200258534-A2.

XX

PD 01-AUG-2002.

XX

PF 16-NOV-2001; 2001WO-US043704.

XX

PR 20-NOV-2000; 2000US-0252222P.

PR 06-FEB-2001; 2001US-0267011P.

PR 28-MAR-2001; 2001US-0279670P.

PR 10-JUL-2001; 2001US-0304037P.

XX

PA (CORI-) CORIXA CORP.

XX

PI Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GE;

XX

DR WPI; 2002-608400/65.

DR N-PSDB; ABV89278.

XX

PT New isolated tumor colon polynucleotide and polypeptide, useful for the  
PT diagnosis, prevention and/or treatment of cancer, in particular colon  
PT cancer.

XX

LOCUS AX136129 2863 bp DNA linear PAT 30-MAY-2001  
 DEFINITION Sequence 51 from Patent EP1067182.  
 ACCESSION AX136129  
 VERSION AX136129.1 GI:14272537  
 KEYWORDS .  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 AUTHORS Ota,T., Isogai,T., Nishikawa,T., Kawai,Y., Sugiyama,T. and Hayashi,K.  
 TITLE Secretory protein or membrane protein  
 JOURNAL Patent: EP 1067182-A 51 10-JAN-2001;  
 Helix Research Institute (JP)

FEATURES Location/Qualifiers  
 source 1..2863  
 /organism="Homo sapiens"  
 /mol\_type="unassigned DNA"  
 /db\_xref="taxon:9606"  
 CDS 231..923  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="CAC39707.1"  
 /db\_xref="GI:14272538"  
 /db\_xref="REMTREMBL:CAC39707"  
 /translation="MASLGLQLVGYYILGLLGLLGTLVAMLLPSWKTSSYVGASIVTAV  
 GFSKGLWMECATHSTGITQCDIYSTLLGLPADIQAAQAMMVTSSAIISSLACIISVVGM  
 RCTVFCQESRAKDRVAVAGGVFFILGGLLGFI PVAWNHLHGILRDFYSPLVPDSMKFEI  
 GEALYLGIISSLFSLIAGIILCFSCSSQRNRSNYYDAYQAQPLATRSSPRPGQPPKVK  
 SEFNSYSLTGYV"

#### ORIGIN

Query Match 99.7%; Score 1470.2; DB 6; Length 2863;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1472; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy	1	GAGAGAAGTCAGCCTGGCAGAGAGACTCTGAAATGAGGGATTAGAGGTGTTCAAGGAGCA	60
Db	110	GAGAGAAGTCAGCCTGGCAGAGAGACTCTGAAATGAGGGATTAGAGGTGTTCAAGGAGCA	169
Qy	61	AGAGCTTCAGCCTGAAGACAAGGGAGCAGTCCCTGAAGACGCTTCTACTGAGAGGTCTGC	120
Db	170	AGAGCTTCAGCCTGGAGACAAGGGAGCAGTCCCTGAAGACGCTTCTACTGAGAGGTCTGC	229
Qy	121	CATGGCCTCTCTTGGCCTCCAACCTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGG	180
Db	230	CATGGCCTCTCTTGGCCTCCAACCTTGTGGGCTACATCCTAGGCCTTCTGGGGCTTTTGGG	289
Qy	181	CACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCAT	240
Db	290	CACACTGGTTGCCATGCTGCTCCCCAGCTGGAAAACAAGTTCTTATGTCGGTGCCAGCAT	349
Qy	241	TGTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGG	300
Db	350	TGTGACAGCAGTTGGCTTCTCCAAGGGCCTCTGGATGGAATGTGCCACACACAGCACAGG	409

Qy	301	CATCACCCAGTGTGACATCTATAGCACCTTCTGGGCCTGCCCGCTGACATCCAGGCTGC	360
Db	410	CATCACCCAGTGTGACATCTATAGCACCTTCTGGGCCTGCCCGCTGACATCCAGGCTGC	469
Qy	361	CCAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGT	420
Db	470	CCAGGCCATGATGGTGACATCCAGTGCAATCTCCTCCCTGGCCTGCATTATCTCTGTGGT	529
Qy	421	GGGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGC	480
Db	530	GGGCATGAGATGCACAGTCTTCTGCCAGGAATCCCGAGCCAAAGACAGAGTGGCGGTAGC	589
Qy	481	AGGTGGAGTCTTTTTTCATCCTTGGAGGCCTCCTGGGATTCAATTCCTGTTGCCTGGAATCT	540
Db	590	AGGTGGAGTCTTTTTTCATCCTTGGAGGCCTCCTGGGATTCAATTCCTGTTGCCTGGAATCT	649
Qy	541	TCATGGGATCCTACGGGACTTCTACTCACCCTGGTGCCTGACAGCATGAAATTTGAGAT	600
Db	650	TCATGGGATCCTACGGGACTTCTACTCACCCTGGTGCCTGACAGCATGAAATTTGAGAT	709
Qy	601	TGGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATCAT	660
Db	710	TGGAGAGGCTCTTTACTTGGGCATTATTTCTTCCCTGTTCTCCCTGATAGCTGGAATCAT	769
Qy	661	CCTCTGCTTTTTCCTGCTCATCCCAGAGAAATCGCTCCAACCTACTACGATGCCTACCAAGC	720
Db	770	CCTCTGCTTTTTCCTGCTCATCCCAGAGAAATCGCTCCAACCTACTACGATGCCTACCAAGC	829
Qy	721	CCAACCTCTTGCCACAAGGAGCTCTCCAAGGCCTGGTCAACCTCCCAAAGTCAAGAGTGA	780
Db	830	CCAACCTCTTGCCACAAGGAGCTCTCCAAGGCCTGGTCAACCTCCCAAAGTCAAGAGTGA	889
Qy	781	GTTCAATTCCTACAGCCTGACAGGGTATGTGTGAAGAACCAGGGGCCAGAGCTGGGGGGT	840
Db	890	GTTCAATTCCTACAGCCTGACAGGGTATGTGTGAAGAACCAGGGGCCAGAGCTGGGGGGT	949
Qy	841	GGCTGGGTCTGTGAAAAACAGTGGACAGCACCCCGAGGGCCACAGGTGAGGGACACTACC	900
Db	950	GGCTGGGTCTGTGAAAAACAGTGGACAGCACCCCGAGGGCCACAGGTGAGGGACACTACC	1009
Qy	901	ACTGGATCGTGTGAGAAGGTGCTGCTGAGGATAGACTGACTTTGGCCATTGGATTGAGCA	960
Db	1010	ACTGGATCGTGTGAGAAGGTGCTGCTGAGGATAGACTGACTTTGGCCATTGGATTGAGCA	1069
Qy	961	AAGGCAGAAATGGGGGCTAGTGTAACAGCATGCAGGTTGAATTGCCAAGGATGCTCGCCA	1020
Db	1070	AAGGCAGAAATGGGGGCTAGTGTAACAGCATGCAGGTTGAATTGCCAAGGATGCTCGCCA	1129
Qy	1021	TGCCAGCCTTTCTGTTTTCTCACCTTGCTGCTCCCCTGCCCTAAGTCCCCAACCTCAA	1080
Db	1130	TGCCAGCCTTTCTGTTTTCTCACCTTGCTGCTCCCCTGCCCTAAGTCCCCAACCTCAA	1189
Qy	1081	CTTGAAACCCCATTTCCCTTAAGCCAGGACTCAGAGGATCCCTTTGCCCTCTGGTTTACCT	1140
Db	1190	CTTGAAACCCCATTTCCCTTAAGCCAGGACTCAGAGGATCCCTTTGCCCTCTGGTTTACCT	1249
Qy	1141	GGGACTCCATCCCCAAACCCACTAATCACATCCCACTGACTGACCCTCTGTGATCAAAGA	1200



Db	1250	 GGGACTCCATCCCCAAACCCACTAATCACATCCCACTGACTGACCCTCTGTGATCAAAGA	1309
Qy	1201	CCCTCTCTCTGGCTGAGGTTGGCTCTTAGCTCATTGCTGGGGATGGGAAGGAGAAGCAGT	1260
Db	1310	 CCCTCTCTCTGGCTGAGGTTGGCTCTTAGCTCATTGCTGGGGATGGGAAGGAGAAGCAGT	1369
Qy	1261	GGCTTTTGTGGGCATTGCTCTAACCTACTTCTCAAGCTTCCCTCCAAAGAACTGATTGG	1320
Db	1370	 GGCTTTTGTGGGCATTGCTCTAACCTACTTCTCAAGCTTCCCTCCAAAGAACTGATTGG	1429
Qy	1321	CCCTGGAACCTCCATCCCACTCTTGTTATGACTCCACAGTGTCCAGACTAATTTGTGCAT	1380
Db	1430	 CCCTGGAACCTCCATCCCACTCCTGTTATGACTCCACAGTGTCCAGACTAATTTGTGCAT	1489
Qy	1381	GAACTGAAATAAAACCATCCTACGGTATCCAGGGAACAGAAAGCAGGATGCAGGATGGGA	1440
Db	1490	 GAACTGAAATAAAACCATCCTACGGTATCCAGGGAACAGAAAGCAGGATGCAGGATGGGA	1549
Qy	1441	GGACAGGAAGGCAGCCTGGGACATTTAAAAAAATA	1475
Db	1550	 GGACAGGAAGGCAGCCTGGGACATTTAAAAAAATTA	1584